Universidade Tecnológica Federal do Paraná
INF72A – Computação 2
Engenharia de Controle e Automação e Engenharia Elétrica
Prof. Marcelo de Oliveira Rosa
Prof. Newton Carlos Will
Prof. Roberto Zanetti Freire
Prof. Robson Ribeiro Linhares

Projeto da Disciplina – Aprendizado de Máquina: Algoritmos de Classificação em Linguagem C

Objetivo: a equipe deverá adaptar um algoritmo de classificação binária ao problema (competição) de classificação (previsão) de pessoas que sobreviveram no naufrágio do navio Titanic. A implementação deve ser feita em linguagem C. Dados e descrição do *dataset* disponíveis aqui.

Requisitos:

- Cada equipe irá escolher um algoritmo de classificação diferente.
- Os dados estão separados em dois arquivos: *train.csv* e *test.csv*. Para treinar o modelo só é permitido usar os dados de treinamento. Para testar o modelo, será utilizado os dados de teste e feito o upload na competição, onde é possível ver o resultado do seu modelo.
- Os critérios de avaliação de desempenho dos métodos serão os mesmos para todas as duplas: os 6 critérios apresentados nos próximos slides.
- Para o conjunto de dados de teste, será considerado, somente Acurácia calculada no site da competição. A equipe precisa submeter o resultado para validar o projeto, mostrando o resultado ao professor durante a apresentação.

Critérios de Avaliação de Desempenho do Método:

Serão considerados 6 critérios de avaliação, todos devem ser apresentados e avaliados para o conjunto de dados de treinamento (matriz de confusão e equações 1 a 5):

1. **Matriz de confusão:**
   - TP: *true positives;*
   - FP: *false positives;*
   - TN: *true negatives;*
   - FN: *false negatives.*

		Predicted class: P	Predicted class: N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

2. **Erro de Previsão (*Prediction Error* - Err):**

$$Err = \frac{FP + FN}{FP + FN + TP + TN}$$

3. **Acurácia (Accuracy - Acc):**

$$Acc = \frac{TP + TN}{FP + FN + TP + TN}$$

$$Acc = 1 - Err$$

4. **Precisão (*Precision* - Pre):**

$$Pre = \frac{TP}{TP + FP}$$

5. **Recall (Rec):**

$$Rec = \frac{TP}{TP + FN}$$

6. **$F1_score$:**

$$F1score = \frac{2 \times Pre \times Rec}{Pre + Rec}$$

Entrega:

1. Código computacional que treina e avalia o método escolhido nos 6 critérios de avaliação de desempenho;
2. Apresentação em slides (10 min) contendo:
   - **Explicação do método (como ele funciona);**

     Incluir autor e ano em que o método foi proposto.

     Explicar como o método funciona.

   - **Procedimentos de treinamento e validação do método;**

     Explicar como é feita a entrada de dados no método.

     Valores de $Err, Acc, Pre, Rec, F1score$ para conjunto de dados de treinamento.

   - **Resultados (critérios de avaliação);**

     Valores de $Err, Acc, Pre, Rec, F1score$ para o conjunto de dados de validação.

     Incluir alguns detalhes de implementação.

   - **Discussão.**

     Comentar se o método foi eficiente, como pode melhorar, o que foi feito de errado ou certo durante a implementação/treinamento/validação.

Universidade Tecnológica Federal do Paraná
INF72A – Computação 2
Engenharia de Controle e Automação e Engenharia Elétrica
Prof. Marcelo de Oliveira Rosa
Prof. Newton Carlos Will
Prof. Roberto Zanetti Freire
Prof. Robson Ribeiro Linhares

Observações:

- A equipe que procurar formas de melhorar a resposta dos métodos e que apresentar o melhor resultado, terá a maior nota;
- Mais informações sobre critérios de desempate serão mencionados em sala de aula.